On the unbiasedness of iterated GLS estimators

Don, F.J.H.; Magnus, J.R.

Published in:
Communications in statistics: Part A: Theory and methods

Publication date:
1980

Citation for published version (APA):
Don, F. J. H., & Magnus, J. R. (1980). On the unbiasedness of iterated GLS estimators. *Communications in statistics: Part A: Theory and methods*, *A9(5)*, 519-527.

COMMUN. STATIST.-THEOR. METH., A9(5), 519-527 (1980)

ON THE UNBIASEDNESS OF ITERATED GLS ESTIMATORS

F.J. Henk Don and Jan R. Magnus

Central Planning Bureau, The Hague

and

University of British Columbia, Vancouver

Key Words & Phrases: generalized least squares; maximum likelihood; iterative estimation; existence of moments; unbiasedness.

ABSTRACT

We formulate sufficient conditions for the existence of the expectation of the iterated generalized least squares estimator, which consequently guarantee its unbiasedness. The analysis is applied to the maximum likelihood estimator in the general linear model with normal disturbances, where a set of assumptions ensures convergence of the iteration as well as unbiasedness.

1. INTRODUCTION

In 1967 Kakwani showed that any estimator which has a sampling error of the form $H(\varepsilon)\cdot\varepsilon$ is unbiased, provided that the estimator has a finite expectation, ε is distributed symmetricly about zero, and H is even (i.e. invariant under change of sign of all elements in ε: $H(\varepsilon) = H(-\varepsilon)$).This result was subsequently used to establish that Zellner's two-step estimator for seemingly unrelated regressions, and Theil's mixed regression estimator are unbiased, if the expectations exist (Kakwani (1967, 1968)). Swamy and Mehta (1969) proved the existence of the expectation of the mixed regression estimator under normality, thus establishing its unbiasedness. In 1973 Fuller and Battese provided an elegant derivation of conditions under which the expectation of the GLS estimator in a variance components model exists, and they proved unbiasedness under normality. Mehta and Swamy (1976) studied a variant of Zellners's two-step estimator and they established its unbiasedness under normality. Finally, Harvey (1978) studied the unbiasedness of L_p-norm and M-estimators.

In this paper we shall consider the iterated GLS estimators

$$b_j = (X'\Omega_j^{-1}X)^{-1} X'\Omega_j^{-1} y \qquad (j=0,1,\ldots) ,$$

where Ω_j is estimated as some function of the residuals from the preceding step. Proposition 1 states that Kakwani's result holds at each step of the iteration. In the second proposition we formulate sufficient conditions for the existence of Eb_j. The two propositions combined say that $Eb_j = \beta$, if ε is symmetric about zero, Ω_j is an even function of the residuals from the preceding step, and $E(\mathrm{tr}\Omega_j)(\mathrm{tr}\Omega_j^{-1})$ exists.
Next, we assume normal disturbances. It is then natural to ask whether the maximum likelihood estimator $\hat{\beta}$ is unbiased. By strengthening an assumption made by Oberhofer and Kmenta (1974), we can state a condition which guarantees not only unbiasedness of $\hat{\beta}$, but also convergence of the iteration to a solution of the maximum likelihood equations. Finally, an adjusted iterative procedure is proposed which may be used in case this condition is violated.

2. TWO USEFUL PROPOSITIONS

We shall consider the linear regression model

$$y = X\beta + \varepsilon ,$$

where y is an (n,1) vector of observations on the dependent variable, X is an (n,k) matrix of the values of the regressors, β is a (k,1) vector of regression coefficients, and ε is an (n,1) disturbance vector. We make the following assumptions:

Assumption 1: $E\varepsilon = 0$, $E\varepsilon\varepsilon' = \Omega$, where Ω is a positive definite matrix whose elements are differentiable functions of a finite and constant number of parameters $\theta_1 \dots \theta_m$.

Assumption 2: X is a fixed matrix of full rank and $n>k$.

Assumption 3: The parameters in β are unrelated to those in θ, where $\theta = (\theta_1 \dots \theta_m)'$.

In order to estimate the parameter vectors β and θ, we choose some starting value θ_0 for θ, and apply an iteration scheme whose jth step ($j=0,1,\dots$) is defined by

$$\begin{aligned}
\Omega_j^{-1} &= \Omega^{-1}(\theta_j) \ , \\
b_j &= (X'\Omega_j^{-1}X)^{-1}X'\Omega_j^{-1}y \ , \\
e_j &= y - Xb_j \ , \\
\theta_{j+1} &= \theta(e_j) \ .
\end{aligned}$$

Thus b_j is estimated by GLS based upon the estimated θ_j, which is estimated as a function - for the moment unspecified - of the residual vector e_{j-1}.

Proposition 1. If ε is symmetric about zero and θ_j is estimated as an even function of the residuals e_{j-1}, then b_j is symmetric about β at *each* step of the iteration.

proof[1]. At the j-th step we have

$$\begin{aligned}
e_j &= y - Xb_j = y - X(X'\Omega_j^{-1}X)^{-1}X'\Omega_j^{-1}y \\
&= \left[I - X(X'\Omega_j^{-1}X)^{-1}X'\Omega_j^{-1}\right]\varepsilon \qquad (j=0,1,\dots) \ .
\end{aligned}$$

Following Kakwani's (1967) approach we see that, if ε changes sign, e_0 will change sign, but θ_1 - being an even function of e_0 - will not be affected. Ω_1 will not be affected either. Thus Ω_1 is an even function of ε. This implies that b_1 is symmetric about β. It also implies that e_1 changes sign if ε changes sign. Since θ_2 is an even function of e_1, it is an even function of ε. Therefore b_2 is symmetric about β. Clearly, the argument can be repeated for $j=3,4,\dots$. Q.E.D.

---

[1] A similar proposition in the context of maximum likelihood is proved in Magnus (1978 lemma 3 and 4).

In other words, b_j is unbiased *if its mean exists*. Sufficient conditions for the existence of Eb_j are the subject of Proposition 2. First we will state the following lemma which will prove useful in the sequel.

Lemma 1. Let Ω be some positive semi-definite matrix of random variables, then the following statements are equivalent:

(a) $E\Omega$ exists

(b) $E\mathrm{tr}\Omega$ exists

(c) $E\mu(\Omega)$ exists,

where $\mu(\Omega)$ denotes the maximal characteristic root of Ω.

proof. We shall establish that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). Now, (a) $\Rightarrow$ (b) is trivial. Further, $\mu(\Omega) \leqq \mathrm{tr}\Omega$ implies that (b) $\Rightarrow$ (c). Finally, since $\mu(\Omega)$ is the maximal characteristic root of Ω, we have $x'\Omega x \leqq \mu(\Omega)x'x$ for any real n-vector x (n being the order of Ω). Let i_h be the h-th column of the identity matrix I_n. Then $i_h'\Omega i_h \leqq \mu(\Omega)i_h'i_h$, i.e. $\omega_{hh} \leqq \mu(\Omega)$. Also for $h \neq k$, $(i_h+i_k)'\Omega(i_h+i_k) \leqq \mu(\Omega)(i_h+i_k)'(i_h+i_k)$, i.e. $\omega_{hh}+\omega_{kk}+2\omega_{hk} \leqq 2\mu(\Omega)$. Obviously $\omega_{hh} \geqq 0$, $\omega_{kk} \geqq 0$. Therefore $\omega_{hk} \leqq \mu(\Omega)$ and (c) $\Rightarrow$ (a).

Q.E.D.

The following proposition can now be proved.

Proposition 2. A sufficient condition for the existence of Eb_j is the existence of

(i) $E\mathrm{tr}(\Omega_j)\mathrm{tr}(\Omega_j^{-1})$.

If ε has fourth moments, then any one of the following two conditions is sufficient too:

(ii) $E\mathrm{tr}(\Omega_j)$ and $E(\mathrm{tr}\Omega_j^{-2})$ exist ,

(iii) $E\mathrm{tr}(\Omega_j^2)$ and $E(\mathrm{tr}\Omega_j^{-1})$ exist .

proof. The sampling error of b_j has the form

$$b_j - \beta = (X'\Omega_j^{-1}X)^{-1}X'\Omega_j^{-1}\varepsilon .$$

Since X has full rank, it is necessary and sufficient to prove the existence of $EH_j\varepsilon$, where

$$H_j = X(X'\Omega_j^{-1}X)^{-1}X\,\Omega_j^{-1}$$

is an idempotent matrix, but *not* symmetric. This is important since it implies that the elements of H_j may not be bounded by some function of X.[2)] We therefore introduce the idempotent and symmetric matrix P_j defined as

$$P_j = \Omega_j^{-\frac{1}{2}}X(X'\Omega_j^{-1}X)^{-1}X'\Omega_j^{-\frac{1}{2}} .$$

Let a be an arbitrary nonstochastic real n-vector. Since $H_j = \Omega_j^{\frac{1}{2}}P_j\Omega_j^{-\frac{1}{2}}$, a necessary and sufficient condition that Eb_j exists is that

$$|a'\Omega_j^{\frac{1}{2}}P_j\Omega_j^{-\frac{1}{2}}\varepsilon|$$

has a finite expectation. Proceeding now as in Fuller and Battese (1973, p. 628-9), we can show that the following inequalities hold:

$$\begin{aligned}|a'\Omega_j^{\frac{1}{2}}P_j\Omega_j^{-\frac{1}{2}}\varepsilon| &\leqq (a'\Omega_j a)^{\frac{1}{2}}\,(\varepsilon'\Omega_j^{-\frac{1}{2}}P_j\Omega_j^{-\frac{1}{2}}\varepsilon)^{\frac{1}{2}}\\ &\leqq (a'\Omega_j a)^{\frac{1}{2}}\,(\varepsilon'\Omega_j^{-1}\varepsilon)^{\frac{1}{2}} \leqq \left[\mu(\Omega_j)a'a\right]^{\frac{1}{2}}\left[\mu(\Omega_j^{-1})\varepsilon'\varepsilon\right]^{\frac{1}{2}}\\ &= (a'a)^{\frac{1}{2}}\left[\mu(\Omega_j)\mu(\Omega_j^{-1})\varepsilon'\varepsilon\right]^{\frac{1}{2}} .\end{aligned}$$

The matrix P_j being idempotent and symmetric, the first inequality is Cauchy-Schwarz'. Moreover, $\mu(P_j) = 1$ and thus the second and third inequality hold since the Rayleigh quotient $x'\Omega x/x'x$ attains its maximum at $\mu(\Omega)$.

Thus, if $E[\mu(\Omega_j)\mu(\Omega_j^{-1})\varepsilon'\varepsilon]^{\frac{1}{2}}$ exists, then Eb_j exists.

Now, since for any two random variables x,y : $(Exy)^2 \leqq (Ex^2)(Ey^2)$, we have for any two nonnegative random variables a,b : $E(ab)^{\frac{1}{2}} \leqq (Ea)^{\frac{1}{2}}(Eb)^{\frac{1}{2}}$, and thus

(i) $E[\mu(\Omega_j)\mu(\Omega_j^{-1})\varepsilon'\varepsilon]^{\frac{1}{2}} \leqq [E\mu(\Omega_j)\mu(\Omega_j^{-1})]^{\frac{1}{2}}\,[tr(\Omega)]^{\frac{1}{2}}$,

(ii) $E[\mu(\Omega_j)\mu(\Omega_j^{-1})\varepsilon'\varepsilon]^{\frac{1}{2}} \leqq [E\mu(\Omega_j)]^{\frac{1}{2}}[E\mu^2(\Omega_j^{-1})]^{\frac{1}{4}}\left[E(\varepsilon'\varepsilon)^2\right]^{\frac{1}{4}}$,

(iii) $E[\mu(\Omega_j)\mu(\Omega_j^{-1})\varepsilon'\varepsilon]^{\frac{1}{2}} \leqq [E\mu(\Omega_j^{-1})]^{\frac{1}{2}}[E\mu^2(\Omega_j)]^{\frac{1}{4}}\left[E(\varepsilon'\varepsilon)^2\right]^{\frac{1}{4}}$.

The proposition now follows from Lemma 1 and the fact that $\mu^2(\Omega) = \mu(\Omega^2)$. Q.E.D.

---

2) If $H_j = (h_{st})$ were bounded by, say, $|h_{st}| \leqq \phi(X)$, then $tr(H_j'H_j)$ would be bounded by $n^2\phi^2(X)$. Also $\varepsilon'H_j'H_j\varepsilon \leqq (trH_j'H_j)\varepsilon'\varepsilon \leqq n^2\phi^2(X)\varepsilon'\varepsilon$, which has a finite expectation. Then $H_j\varepsilon$ has a finite expectation. Of course, when $M = (m_{ij})$ is idempotent *and* symmetric, then $|m_{ij}| \leqq 1$.

Note that Theorem 2 in Fuller and Battese (1973) is a straightforward application of our Proposition 2 (iii).

Summarizing this section, we have established that b_j is an unbiased estimator for β at each step of the iteration, provided only that ε is symmetric about zero, θ_j is an even function of e_{j-1}, and $E(\mathrm{tr}\Omega_j)(\mathrm{tr}\Omega_j^{-1})$ exists.

3. THE UNBIASEDNESS OF THE ML ESTIMATOR FOR β

Let us now turn to the maximum likelihood (ML) estimation of β and θ, and investigate the condition under which the ML estimator $\hat{\beta}$ is unbiased. We need one more assumption.

Assumption 4: ε is normally distributed.

The probability density of y is

$$(2\pi)^{-\frac{n}{2}}|\Omega|^{-\frac{1}{2}}\exp-\tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon .$$

Thus, the loglikelihood takes the form

$$\Lambda(\beta,\theta) = \gamma + \tfrac{1}{2}\log|\Omega^{-1}| - \tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon ,$$

where γ is a constant.

As shown in Magnus (1978, Theorem 1), the ML equations are

$$\text{(i)} \quad \hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

$$\text{(ii)} \quad \mathrm{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta^{(h)}}\Omega\right)_{\theta=\hat{\theta}} = e'\left(\frac{\partial\Omega^{-1}}{\partial\theta^{(h)}}\right)_{\theta=\hat{\theta}}e \qquad (h=1\ldots m) ,$$

where $\theta^{(h)}$ is the h-th component of θ, $e = y-X\hat{\beta}$, and $\hat{\Omega} = \Omega(\hat{\theta})$.

The ML estimates are obtained by using the iterative scheme of section 2. This means that θ_{j+1} is estimated as the solution of

$$\mathrm{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta^{(h)}}\Omega\right) = e_j'\frac{\partial\Omega^{-1}}{\partial\theta^{(h)}}e_j \qquad (h=1\ldots m) .$$

Now, even though it is often impossible to express θ_{j+1} explicitly in terms of e_j, it is clear that θ_{j+1} is an even function of e_j. Furthermore ε is symmetric about zero. Thus, b_j is unbiased if $E\mathrm{tr}(\Omega_j)\mathrm{tr}(\Omega_j^{-1})$ exists.

When considering the ML estimator $\hat{\beta}$, we must make sure that the iterative procedure in fact converges to a solution of the ML equations. This problem has been studied by Oberhofer and Kmenta

(1974). We shall need one last assumption.

Assumption 5: There exist fixed numbers $\lambda > 0$, $M > 0$ such that when $U := \{\theta \mid \lambda \leq |\Omega| ; \theta'\theta \leq M\}$ the following is true: with probability one a number s may be found such that for any $\beta^* \in \mathbb{R}^k$, $\theta^* \in U$ we have

$$\Lambda(\beta^*,\theta^*) \geq s \Rightarrow \hat{\theta}(\beta^*) \in U$$

where $\hat{\theta}(\beta^*)$ denotes the value of θ that maximizes $\Lambda(\beta^*,\theta)$.

The above Assumption 5 is stronger than the corresponding Assumption 6 in Oberhofer and Kmenta (1974, p. 583). Thus, while Oberhofer and Kmenta allow the values of λ and M to change for different samples, we need the statement to hold with fixed λ and M for (almost) all possible samples.

The Assumptions (1) - (5) are sufficient to guarantee that the iterative procedure converges to a solution of the ML equations (Oberhofer and Kmenta (1974, Theorem 1)), but not only that. If Assumption 5 is satisfied, fixed numbers ℓ_1 and ℓ_2 exist such that almost surely

$$0 < \ell_1 \leq \lambda(\hat{\Omega}) \leq \ell_2 ,$$

where $\lambda(\hat{\Omega})$ stands for any eigenvalue of the estimated covariance matrix $\Omega(\hat{\theta})$.[3] This implies that almost surely

$$\mu(\hat{\Omega})\ \mu(\hat{\Omega}^{-1}) \leq \ell_2/\ell_1 .$$

Thus $\mu(\hat{\Omega})\mu(\hat{\Omega}^{-1})$ is essentially bounded. Then $tr(\hat{\Omega})tr(\hat{\Omega}^{-1})$ is also essentially bounded. Hence its expectation exists. Summarizing, we have proved the following

Proposition 3. Under the Assumptions (1) - (5), b_j is an unbiased estimator of β at each step of the iteration. The iterative procedure converges to a solution of the ML equations and the resulting ML estimator $\hat{\beta}$ is unbiased too.

It should be noted that Proposition 3 rests crucially on Assumption 5 which is often difficult, if at all possible, to verify. In many applications, possible violation of Assumption 5 may be remedied by the introduction of some adjusted procedure like the one

[3] Oberhofer and Kmenta (1974, p. 583).

recently proposed by Sargan (1978). Thus, a procedure could be defined which differs from the iterative scheme of section 2 in that the covariance matrix $\hat{\Omega}_j$ is adjusted at each step of the iteration in such a way that the ratio of its largest to its smallest eigenvalue remains bounded by some a priori fixed value. The resulting estimates may of course differ from the ML estimates, but by setting the a priori bound high enough, this is unlikely to occur in practice. Moreover, the adjusted iteration procedure will lead to an unbiased estimate of β.

Example: Autocorrelated errors

Let $\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t$, $E\xi = 0$, $E\xi\xi' = \sigma^2 I_n$, $t=1\ldots n$, with $\xi = (\xi_1 \ldots \xi_n)'$.

On the assumption that the process is stationary, an appropriate estimation procedure is maximum likelihood with ε_0 specified through

$$\varepsilon_1 - \rho\varepsilon_0 = \sqrt{1-\rho^2}\ \varepsilon_1 .^{4)}$$

The covariance matrix of the disturbance vector $\varepsilon = (\varepsilon_1 \ldots \varepsilon_n)'$ is

$$\Omega = E\varepsilon\varepsilon' = \sigma^2 (A'A)^{-1} ,$$

with

$$A = \begin{bmatrix} \sqrt{1-\rho^2} & & & 0 \\ -\rho & 1 & & \\ & \ddots & \ddots & \\ 0 & & -\rho & 1 \end{bmatrix} .$$

One verifies that $\mathrm{tr}(A'A) = n + (n-2)\rho^2$ and $\mathrm{tr}(A'A)^{-1} = n/(1-\rho^2)$. If $n \geq 2$ and the process is stationary (i.e. $\rho^2 < 1$), we have

$$\frac{n(n-1)}{1-\rho^2} < (\mathrm{tr}\Omega)(\mathrm{tr}\Omega^{-1}) \leq \frac{2n(n-1)}{1-\rho^2} .$$

Hence, the existence of $E(\mathrm{tr}\hat{\Omega}_j)(\mathrm{tr}\hat{\Omega}_j^{-1})$ depends upon the existence of $E\,\dfrac{1}{1-\hat{\rho}^2}$.

Though the iterative procedure guarantees that $\hat{\rho}_j^2 < 1$ (in contrast with the Cochrane-Orcutt procedure, see Beach and MacKinnon (1978)), Assumption 5 is not necessarily satisfied. Thus, to ensure unbiasedness of b_j, we may impose $\rho_j^2 < 1 - \alpha$ for some a priori chosen $\alpha > 0$.

4) See Magnus (1978) or Beach and MacKinnon (1978) for further details.

The resulting estimator of β will be unbiased and the estimates will equal the ML estimates in all cases where the bound restriction is not violated during the iteration.

BIBLIOGRAPHY

Beach, C.M. and J.G. MacKinnon (1978), "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors", *Econometrica 46*, 51-58.

Fuller, W.A. and G.E. Battese (1973), "Transformations for Estimation of Linear Models with Nested-Error Structure", *J. Amer. Statist. Assoc. 68*, 626-32.

Harvey, A.C. (1978), "On the unbiasedness of robust regression estimators", *Commun. Statist. 7*, Issue 8.

Kakwani, N.C. (1967), "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators", *J. Amer. Statist. Assoc. 62*, 141-2.

Kakwani, N.C. (1968), "Note on the Unbiasedness of a Mixed Regression Estimator", *Econometrica 36*, 610-1.

Magnus, J.R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", *J. Econometrics 7*, 218-312.

Mehta, J.S. and P.A.V.B. Swamy (1976), "Further Evidence on the Relative Efficiencies of Zellner's Seemingly Unrelated Regressions Estimators",*J. Amer. Statist. Assoc. 71*, 634-9.

Oberhofer, W. and J. Kmenta (1974), "A General Procedure for obtaining Maximum Likelihood Estimates in Generalized Regression Models", *Econometrica 42*,579-90.

Sargan, J.D. (1978), "On the Existence of the Moments of 3SLS Estimators", *Econometrica 46*, 1329-1350.

Swamy, P.A.V.B. and J.S. Mehta (1969), "On Theil's Mixed Regression Estimator", *J. Amer. Statist. Assoc. 64*, 273-6.

Received May, 1979; Revised October, 1979.

Refereed by P. A. V. B. Swamy, Federal Reserve Board, Washington, D.C.